APPENDIX A – SYSTEM OUTPUT VALUE MODELS

EDR SCORING

The EDR value score for a system is defined to be the sum of the values of all of the system's output entity tokens, normalized by the sum of the values of all reference entity tokens. The maximum possible EDR value score is 100 percent.

$$EDR\_Value_{sys} = \sum_i value\_of\_sys\_token_i \Big/ \sum_j value\_of\_ref\_token_j$$

The value of each system token is based on its attributes and on how well it matches its corresponding reference token. A globally optimum correspondence between system and reference tokens which maximizes *EDR_Value* is determined and used, subject to the constraint of one-to-one mapping between system and reference tokens.[15] The value of a system token is defined to be the difference of two value terms, one that is in accord with the reference token and one that is not. In this formula, *Mentions_Value*(*sys*,*ref*) and *Mentions_Value*(*sys*,$\overline{ref}$) measure the value of system mentions that correspond and that don't correspond to reference mentions, respectively.

$$\begin{aligned} Value(sys,ref) &= Element\_Value(sys,ref) \cdot Mentions\_Value(sys,ref) \\ &\quad - W_{FA} \cdot Element\_Value(sys) \cdot Mentions\_Value(sys,\overline{ref}) \end{aligned}$$

Element_Value is a function of the attributes of the system token and, if mapped, how well they match those of the corresponding reference token. In particular, *Element_Value* is defined as the product of the values of the token's attributes, specifically the token's ***type***, ***subtype*** and ***class***. This value is then reduced for any attribute errors for the attributes ***type***, ***subtype*** and ***class***, using the attribute error weighting parameters, $\{W_{err\text{-}attribute}\}$.

$$Element\_Value(sys,ref) = \min\left(\prod_{\substack{attribute= \\ type,subtype,class}} AttrValue(attribute_{sys}), \prod_{\substack{attribute= \\ type,subtype,class}} AttrValue(attribute_{ref}) \right) \cdot \prod_{\substack{attribute= \\ type,subtype,class}} W_{err-attribute}$$

$$Element\_Value(sys) = \prod_{\substack{attribute= \\ type,subtype,class}} AttrValue(attribute_{sys})$$

Mentions_Value is a function of the mutual mention value (*MMV*) between the mentions of the system token and, if mapped, those of the corresponding reference token. A mention's *MMV* depends on the mention's ***type*** value parameter, *MTypeValue*, with this value being reduced for any errors in the mention attributes ***type***, ***role*** and ***style***, using the mention attribute error weighting parameters, $\{ W_{Merr} \}$.

$$MMV(mention_{sys}, mention_{ref}) = \begin{cases} \min\left(MTypeValue(mention_{sys}), MTypeValue(mention_{ref}) \right) \cdot \prod_{\substack{attribute= \\ type,role,style}} W_{Merr-attribute} & \text{if } mention_{sys} \text{ and } mention_{ref} \text{ correspond} \\ 0 & \text{otherwise} \end{cases}$$

$$MMV(mention_{sys}) = MTypeValue(mention_{sys})$$

For each pairing of a system token with a reference token, an optimum correspondence between the mentions of the system and reference tokens is determined. This mapping maximizes *Mentions_Value*, subject to the constraint of one-to-one mapping between system and reference mentions.

Mentions_Value is computed using one of two formulas, depending on whether valuation is ***mention***-weighted or ***level***-weighted. For mention-weighted valuation *Mentions_Value* is simply the sum of *MMV* over all mentions in all documents. For level-weighted valuation *Mentions_Value* is determined by a system token's ***level***[16] (and the level of its corresponding reference token), by the degree of correspondence between system and reference mentions, and by the number of documents in which the token is mentioned. Note the use of W_{CR}, which is a weighting factor used to reduce the penalty for making co-reference errors. W_{CR} de-weights the MMV for those false alarm mentions that are legitimate mentions of some other entity.

---

[15] System tokens and reference tokens are permitted to correspond only if they each have at least one mention in correspondence.

[16] A token's ***level*** is the highest (i.e., the most valued) type of mention that is used to refer to that token. Thus, for example, the level of a token is NAM (named) if any one of its mentions is of type NAM, because NAM mentions are more valuable than NOM mentions. If none of its mentions is of type NAM, but at least one mention is of type NOM, then the "level" of that token would be NOM (nominal). The level attribute value is equal to the mention type value for that level: *AttrValue*(*level*) = *MTypeValue*(*level*).

For mention-weighted scoring, *Mentions_Value* is:

$$Mentions\_Value(sys, ref) = \sum_{\substack{all \\ docs}} \left(\sum_{\substack{all\ sys\ mentions \\ in\ doc\ that\ map \\ to\ ref\ mentions}} MMV(mention_{sys}, mention_{ref}) \right)$$

$$Mentions\_Value(sys, \overline{ref}) = \sum_{\substack{all \\ docs}} \left(\sum_{\substack{all\ sys\ mentions \\ in\ doc\ that\ \textbf{don't} \\ map\ to\ ref\ mentions}} W_{CR}(mention_{sys}) \cdot MMV(mention_{sys}) \right)$$

For level-weighted scoring, *Mentions_Value* is:

$$Mentions\_Value(sys, ref) = \sum_{\substack{all \\ docs}} \left(MTypeValue(level_{doc,ref}) \cdot \sum_{\substack{all\ sys\ mentions \\ in\ doc\ that\ map \\ to\ ref\ mentions}} MMV(ment._{sys}, ment._{ref}) \Big/ \sum_{\substack{all\ ref\ mentions \\ in\ doc}} MMV(ment._{ref}) \right)$$

$$Mentions\_Value(sys, \overline{ref}) = \sum_{\substack{all \\ docs}} \left(MTypeValue(level_{doc,sys}) \cdot \sum_{\substack{all\ sys\ mentions \\ in\ doc\ that\ \textbf{don't} \\ map\ to\ ref\ mentions}} W_{CR}(ment._{sys}) \cdot MMV(ment._{sys}) \Big/ \sum_{\substack{all\ sys\ mentions \\ in\ doc}} MMV(ment._{sys}) \right)$$

System mentions and reference mentions are permitted to correspond only if their **heads** have a mutual overlap of at least *min_overlap* and the text of their **heads** share a (fractional) consecutive string of characters[17] of at least *min_text_match*. Mention regions and overlaps are measured in terms of *number of characters* for text input, in terms of *time* for audio input, and in terms of *area* for image input.

$$mutual\_overlap = \frac{sys\_head \cap ref\_head}{\max(sys\_head, ref\_head)}$$

$$fractional\_consecutive\_string = \frac{\left(\begin{array}{c} \#\ of\ characters\ in\ the\ longest\ consecutive\ string\ of\ characters \\ that\ is\ contained\ in\ both\ system\ and\ reference\ mention\ head\ texts \end{array} \right)}{\max\left(\begin{array}{l} \#\ of\ characters\ in\ system\ mention\ head\ text, \\ \#\ of\ characters\ in\ reference\ mention\ head\ text \end{array} \right)}$$

The current default scoring parameters for EDR are given in Table 14.

Table 14 Default parameters for scoring EDR performance

***Element_Value* parameters**				***Mentions_Value* parameters**			
Attribute	$W_{err\text{-}attribute}$	**Attribute Value**	***AttrValue***	**Attribute**	$W_{Merr\text{-}attribute}$	**Attribute Value**	***MTypeValue***
Type	0.50	(all types)	1.00	Type	0.90	NAM	1.00
Class	0.75	SPC	1.00			NOM	0.50
		(not SPC)	0.00			PRO	0.10
Subtype	0.90	n/a	n/a	Role	0.90	n/a	n/a
Valuation = level-weighted		W_{FA} = 0.75		Style	0.90	n/a	n/a
				min_overlap = 0.30		***min_text_match*** = 0.00	
				W_{CR} = 1.0 for legitimate reference mentions *(= 1.0 otherwise)*			

[17] This requirement of a common substring in both system and output mention heads was invoked to account for errors in transcribing speech and image data into text. The intent is to require a mention be meaningful and relevant in order to be counted.

VAL SCORING

The VAL value score for a system is defined to be the sum of the values of all of the system's output value tokens, normalized by the sum of the values of all of the reference value tokens. The maximum possible VAL value score is 100 percent.

$$VAL\_Value_{sys} = \sum_i value\_of\_sys\_token_i \Big/ \sum_j value\_of\_ref\_token_j$$

The value of each system token is based on its attributes and on how well it matches its corresponding reference token. A globally optimum correspondence between system and reference tokens which maximizes *VAL_Value* is determined and used, subject to the constraint of one-to-one mapping between system and reference tokens.[15] The value of a system token is defined as the product of two factors that represent both the inherent value of the token and how accurately the token's attributes are recognized and the token's mentions are detected.

$$\begin{aligned} Value(sys,ref) &= Element\_Value(sys,ref)\cdot Mentions\_Value(sys,ref) \\ &\quad - W_{FA}\cdot Element\_Value(sys)\cdot Mentions\_Value(sys,\overline{ref}) \end{aligned}$$

In this formula, *Mentions_Value*(*sys*,*ref*) and *Mentions_Value*(*sys*, $\overline{ref}$) measure the value of system mentions that correspond and that don't correspond to reference mentions, respectively.

Element_Value is a function of the attributes of the system token and, if mapped, how well they match those of the corresponding reference token. In particular, *Element_Value* is defined as the product of the values of the token's attributes, specifically the token's ***type*** and ***subtype***. This value is then reduced for any attribute errors for the attributes ***type*** and ***subtype***, using the attribute error weighting parameters, $\{W_{err\text{-}attribute}\}$.

$$Element\_Value(sys,ref) = \min\left(\prod_{\substack{attribute=\\ type,subtype}} AttrValue(attribute_{sys}), \prod_{\substack{attribute=\\ type,subtype}} AttrValue(attribute_{ref})\right) \cdot \prod_{\substack{attribute=\\ type,subtype}} W_{err-attribute}$$

$$Element\_Value(sys) = \prod_{\substack{attribute=\\ type,subtype}} AttrValue(attribute_{sys})$$

Mentions_Value is a function of the mutual mention value (*MMV*) between the mentions of the system token and, if mapped, those of the corresponding reference token. A mention's *MMV* is simply the value 1 if the mentions correspond, 0 otherwise.

$$MMV(mention_{sys}, mention_{ref}) = \begin{cases} 1 & \text{if } mention_{sys} \text{ and } mention_{ref} \text{ correspond} \\ 0 & \text{otherwise} \end{cases}$$

$$MMV(mention_{sys}) = 1$$

For each pairing of a system token with a reference token, an optimum correspondence between the mentions of the system and reference tokens is determined. This mapping maximizes *Mentions_Value*, subject to the constraint of one-to-one mapping between system and reference mentions.

Mentions_Value is computed using one of two formulas, depending on whether valuation is ***mention***-weighted or ***level***-weighted. For mention-weighted valuation *Mentions_Value* is simply the sum of *MMV* over all mentions in all documents. For level-weighted valuation *Mentions_Value* is determined by the degree of correspondence between system and reference mentions and by the number of documents in which the token is mentioned. Note the use of W_{CR}, which is a weighting factor used to reduce the penalty for making co-reference errors. W_{CR} de-weights the MMV for those false alarm mentions that are legitimate mentions of some other entity.

For mention-weighted scoring, *Mentions_Value* is:

$$Mentions\_Value(sys, ref) = \sum_{\substack{all \\ docs}} \left(\sum_{\substack{all\ sys\ mentions \\ in\ doc\ that\ map \\ to\ ref\ mentions}} MMV(mention_{sys}, mention_{ref}) \right)$$

$$Mentions\_Value(sys, \overline{ref}) = \sum_{\substack{all \\ docs}} \left(\sum_{\substack{all\ sys\ mentions \\ in\ doc\ that\ \textbf{don't} \\ map\ to\ ref\ mentions}} W_{CR}(mention_{sys}) \cdot MMV(mention_{sys}) \right)$$

For level-weighted scoring, *Mentions_Value* is:

$$Mentions\_Value(sys, ref) = \sum_{\substack{all \\ docs}} \left(\sum_{\substack{all\ sys\ mentions \\ in\ doc\ that\ map \\ to\ ref\ mentions}} MMV(mention_{sys}, mention_{ref}) \Big/ \sum_{\substack{all\ ref\ mentions \\ in\ doc}} MMV(mention_{ref}) \right)$$

$$Mentions\_Value(sys, \overline{ref}) = \sum_{\substack{all \\ docs}} \left(\sum_{\substack{all\ sys\ mentions \\ in\ doc\ that\ \textbf{don't} \\ map\ to\ ref\ mentions}} W_{CR}(mention_{sys}) \cdot MMV(mention_{sys}) \Big/ \sum_{\substack{all\ sys\ mentions \\ in\ doc}} MMV(mention_{sys}) \right)$$

System mentions and reference mentions are permitted to correspond only if their **extents** have a mutual overlap of at least *min_overlap*. Mention regions and overlaps are measured in terms of *number of characters* for text input, in terms of *time* for audio input, and in terms of *area* for image input.

$$mutual\_overlap = \frac{sys\_extent \cap ref\_extent}{\max(sys\_extent, ref\_extent)}$$

The current default parameters for VAL scoring are given in Table 15.

Table 15 Default parameters for scoring VAL performance

***Element_Value* parameters**				***Mentions_Value* parameters**	
Attribute	$W_{err\text{-}attribute}$	**Attribute Value**	***AttrValue***	$W_{Merr\text{-}attribute}$	0.90 (for all attributes)
Type	0.50	(all types)	1.00	W_{CR}	1.00
Subtype	0.90	n/a	n/a	***min_overlap***	0.30
$W_{FA} = 0.75$					

TERN SCORING

The TERN value score for a system is defined to be the sum of the values of all of the system's output timex2 tokens, normalized by the sum of the values of all of the reference timex2 tokens. The maximum possible timex2 value score is 100 percent.

$$TERN\_Value_{sys} = \sum_i value\_of\_sys\_token_i \Big/ \sum_j value\_of\_ref\_token_j$$

The value of each system token is based on its attributes and on how well it matches its corresponding reference token. A globally optimum correspondence between system and reference tokens which maximizes *TERN_Value* is determined and used, subject to the constraint of one-to-one mapping between system and reference tokens.[15] The value of a system token is defined as the product of two factors that represent both the inherent value of the token and how accurately the token's attributes are recognized and the token's mentions are detected.

$$\begin{aligned} Value(sys,ref) = {} & Element\_Value(sys,ref) \cdot Mentions\_Value(sys,ref) \\ & - W_{FA} \cdot Element\_Value(sys) \cdot Mentions\_Value(sys,\overline{ref}) \end{aligned}$$

In this formula, *Mentions_Value*(*sys*,*ref*) and *Mentions_Value*(*sys*, $\overline{ref}$) measure the value of system mentions that correspond and that don't correspond to reference mentions, respectively.

Element_Value depends on how well the attributes of the system token match those of the corresponding reference token. The inherent value of a token is defined as a sum of attribute value parameters, *AttrValue*, summed over all attributes which exist and which are the same for both the system and reference tokens. If a system token is unmapped, then the value of that token is weighted by a false alarm penalty, W_{FA}.

$$Element\_Value(sys,ref) = \sum_{\substack{\text{for all existing } sys \text{ attributes in the set} \\ \{type,mod,set,val,anchor\_dir,anchor\_val\}}} \begin{cases} AttrValue(attribute) & \text{if } attribute_{sys} = attribute_{ref} \\ 0 & \text{otherwise} \end{cases}$$

$$Element\_Value(sys) = \sum_{\substack{\text{for all existing } sys \text{ attributes in the set} \\ \{type,mod,set,val,anchor\_dir,anchor\_val\}}} AttrValue(attribute)$$

Mentions_Value is a function of the mutual mention value (*MMV*) between the mentions of the system token and, if mapped, those of the corresponding reference token. A mention's *MMV* is simply the value 1 if the mentions correspond, 0 otherwise.

$$MMV(mention_{sys}, mention_{ref}) = \begin{cases} 1 & \text{if } mention_{sys} \text{ and } mention_{ref} \text{ correspond} \\ 0 & \text{otherwise} \end{cases}$$

$$MMV(mention_{sys}) = 1$$

For each pairing of a system token with a reference token, an optimum correspondence between the mentions of the system and reference tokens is determined. This mapping maximizes *Mentions_Value*, subject to the constraint of one-to-one mapping between system and reference mentions.

Mentions_Value is computed using one of two formulas, depending on whether valuation is ***mention***-weighted or ***level***-weighted. For mention-weighted valuation *Mentions_Value* is simply the sum of *MMV* over all mentions in all documents. For level-weighted valuation *Mentions_Value* is determined by the degree of correspondence between system and reference mentions and by the number of documents in which the token is mentioned. Note the use of W_{CR}, which is a weighting factor used to reduce the penalty for making co-reference errors. W_{CR} de-weights the MMV for those false alarm mentions that are legitimate mentions of some other entity.

For mention-weighted scoring, *Mentions_Value* is:

$$Mentions\_Value(sys, ref) = \sum_{\substack{all \\ docs}} \left(\sum_{\substack{all\ sys\ mentions \\ in\ doc\ that\ map \\ to\ ref\ mentions}} MMV(mention_{sys}, mention_{ref}) \right)$$

$$Mentions\_Value(sys, \overline{ref}) = \sum_{\substack{all \\ docs}} \left(\sum_{\substack{all\ sys\ mentions \\ in\ doc\ that\ \textbf{don't} \\ map\ to\ ref\ mentions}} W_{CR}(mention_{sys}) \cdot MMV(mention_{sys}) \right)$$

For level-weighted scoring, *Mentions_Value* is:

$$Mentions\_Value(sys, ref) = \sum_{\substack{all \\ docs}} \left(\sum_{\substack{all\ sys\ mentions \\ in\ doc\ that\ map \\ to\ ref\ mentions}} MMV(mention_{sys}, mention_{ref}) \Big/ \sum_{\substack{all\ ref\ mentions \\ in\ doc}} MMV(mention_{ref}) \right)$$

$$Mentions\_Value(sys, \overline{ref}) = \sum_{\substack{all \\ docs}} \left(\sum_{\substack{all\ sys\ mentions \\ in\ doc\ that\ \textbf{don't} \\ map\ to\ ref\ mentions}} W_{CR}(mention_{sys}) \cdot MMV(mention_{sys}) \Big/ \sum_{\substack{all\ sys\ mentions \\ in\ doc}} MMV(mention_{sys}) \right)$$

System mentions and reference mentions are permitted to correspond only if their **extents** have a mutual overlap of at least *min_overlap*. Mention regions and overlaps are measured in terms of *number of characters* for text input, in terms of *time* for audio input, and in terms of *area* for image input.

$$mutual\_overlap = \frac{sys\_extent \cap ref\_extent}{\max(sys\_extent, ref\_extent)}$$

The current default parameters for TERN scoring are given in Table 16.

Table 16 Default parameters for scoring TERN performance

***Element_Value* parameters**		***Mentions_Value* parameters**	
W_{FA}	0.75	W_{CR}	1.00
attribute	***AttrValue***	min_overlap	0.30
type	0.10		
anchor_dir	0.25		
anchor_val	0.50		
mod	0.10		
set	0.10		
val	1.00		

RDR SCORING

The RDR value score for a system is defined to be the sum of the values of all of the system's output relation tokens, normalized by the sum of the values of all reference relation tokens. The maximum possible RDR value score is 100 percent.

$$RDR\_Value_{sys} = \sum_i value\_of\_sys\_token_i \Big/ \sum_j value\_of\_ref\_token_j$$

The value of each system token is based on its attributes and arguments and on how well they match those of a corresponding reference token. A globally optimum correspondence between system and reference tokens which maximizes *RDR_Value* is determined and used, subject to the constraint of one-to-one mapping between system and reference tokens. System tokens and reference tokens are permitted to correspond only if they have some nominal basis for correspondence. The required nominal basis is selectable from the set of minimal conditions listed in Table 17.

Table 17 Conditions required for correspondence between system and reference relation tokens

Condition	Description
arguments	At least one argument in the system token must be mappable to an argument in the reference token.
extents	The system and reference tokens must each have at least one mention extent in correspondence with the other.
both	Both the **arguments** condition and the **extents** condition must be met.
either	Either the **arguments** condition or the **extents** condition must be met.
all	All arguments in the reference token must be one-to-one mappable to arguments in the system token.
all+extents	Both the **all** condition and the **extents** condition must be met.

The value of a system token is defined by the following formula:

$$\begin{aligned} Value(sys,ref) &= Element\_Value(sys,ref) \cdot Arguments\_Value(sys,ref) \\ &\quad - W_{FA} \cdot Element\_Value(sys) \cdot Arguments\_Value(sys,\overline{ref}) \end{aligned}$$

In this expression for the *Value* of a system token, *Arguments_Value*(*sys*,*ref*) and *Arguments_Value*(*sys*, $\overline{ref}$) measure the value of system arguments that correspond and that don't correspond to reference arguments, respectively.

Element_Value is a function of the attributes of the system token and, if mapped, how well they match those of the corresponding reference token. The inherent value of a token is defined as the product of the token's attribute value parameters, *AttrValue*, for the attributes ***type*** and ***modality***. This inherent value is reduced for any attribute errors (i.e., for any difference between the values of system and reference attributes), using the error weighting parameters, $\{W_{err\text{-}attribute}\}$. If a system token is unmapped, then the value of that token is weighted by a false alarm penalty, W_{FA}.

$$Element\_Value(sys,ref) = \min\left(\begin{array}{l} \prod_{\substack{attribute= \\ type,modality}} AttrValue(attribute_{sys}), \\ \prod_{\substack{attribute= \\ type,modality}} AttrValue(attribute_{ref}) \end{array} \right) \cdot \prod_{\substack{attribute= \\ type,subtype,modality,tense}} W_{err-attribute}$$

$$Element\_Value(sys) = \prod_{\substack{attribute= \\ type,modality}} AttrValue(attribute_{sys})$$

Arguments_Value is a function of the mutual argument value (*MAV*) between the arguments of the system token and, if mapped, those of the corresponding reference token. An argument's *MAV*, if mapped, is equal to the mapped value of the elements serving as arguments, *Value*(arg_{sys},arg_{ref}), but reduced in value if the system's argument role is in error.

$$MAV(arg_{sys},arg_{ref}) = Value(arg_{sys},arg_{ref}) \cdot W_{err-role} \cdot W_{err-asym} \quad \text{if } Mentions\_Value(arg_{sys},arg_{ref}) > 0$$

$$MAV(arg_{sys}) = Value(arg_{sys},arg_{sys})$$

There are several requirements that must be satisfied in order for a reference argument to be considered to be in correspondence to a system argument. First, note that there are two required arguments, namely the two arguments for which the relation is being asserted. These arguments have roles called "Arg-1" and "Arg-2", and there may be only one Arg-1 and one Arg-2 argument.[18] The requirements for correspondence are listed in Table 18.

Table 18 Conditions required for correspondence between system and reference relation arguments

Condition	Requirement
Always	The reference argument must be mappable to the system argument. That is, they must have at least one mention in correspondence.
Argument role is Arg-1 or Arg-2 and the relation symmetric	The reference argument role may be either "Arg-1" or "Arg-2", and no role mismatch penalty is imposed.
Argument role is Arg-1 or Arg-2 and the relation is **not** symmetric	The reference argument role may be either "Arg-1" or "Arg-2", but an asymmetry error penalty, $W_{err\text{-}asym}$, is imposed.
If the "**mapped**" argument option is invoked	The reference argument must correspond to the system argument. That is, they must be mapped to each other at the argument level.

For each pairing of a system relation token with a reference relation token, an optimum correspondence between system arguments and reference arguments that maximizes *Arguments_Value* is determined and used. This optimum mapping is constrained to be a one-to-one mapping between system and reference arguments.

Arguments_Value is computed using the following formula:

$$Arguments\_Value(sys, ref) = \sum_{\substack{all\ arg_{sys}\ with\ a \\ corresponding\ arg_{ref}}} \left(\sum_{\substack{all\ docs\ that \\ mention\ the\ relation}} MAV_{doc}(arg_{sys}, arg_{ref}) \right)$$

$$Arguments\_Value(sys, \overline{ref}) = \sum_{\substack{all\ arg_{sys}\ with\ \mathbf{no} \\ corresponding\ arg_{ref}}} \left(\sum_{\substack{all\ docs\ that \\ mention\ the\ relation}} MAV_{doc}(arg_{sys}, arg_{sys}) \right)$$

The current default scoring parameters for RDR are given in Table 19.

Table 19 Default parameters for scoring RDR performance

***Element_Value* parameters**			***Arguments_Value* parameters**	
Relation mapping requirements (Table 17) = **"arguments"**			**"mapped"** *arguments optional requirement* **NOT** *invoked* (Table 18)	
W_{FA} = 0.75				
attribute	***AttrValue***	$W_{err\text{-}attribute}$	**Both** Arg-1 and Arg-2 arguments must be mappable (i.e., must have non-null *MAV*'s)	
Type	1.00 for all types	1.00		
Subtype	(not applicable)	0.70	$W_{err\text{-}role}$	0.75
Modality	1.00 for all types	0.75	$W_{err\text{-}asym}$	0.70
Tense	(not applicable)	1.00		

[18] Arg-1 and Arg-2 are the only roles for which the number of arguments is limited.

VDR SCORING

The VDR value score for a system is defined to be the sum of the values of all of the system's output event tokens, normalized by the sum of the values of all reference event tokens. The maximum possible VDR value score is 100 percent.

$$VDR\_Value_{sys} = \sum_i value\_of\_sys\_token_i \Big/ \sum_j value\_of\_ref\_token_j$$

The value of each system token is based on its attributes and arguments and on how well they match those of a corresponding reference token. A globally optimum correspondence between system and reference tokens which maximizes *VDR_Value* is determined and used, subject to the constraint of one-to-one mapping between system and reference tokens. System tokens and reference tokens are permitted to correspond only if they have some nominal basis for correspondence. The required nominal basis is selectable from the set of minimal conditions listed in Table 20. Note that the condition selected applies to both VDR and RDR.

Table 20 Conditions required for correspondence between system and reference event tokens

Condition	Description
arguments	At least one argument in the system token must be mappable to an argument in the reference token.
extents	The system and reference tokens must each have at least one mention extent in correspondence with the other.
both	Both the **arguments** condition and the **extents** condition must be met.
either	Either the **arguments** condition or the **extents** condition must be met.
all	All arguments in the reference token must be one-to-one mappable to arguments in the system token.
all+extents	Both the **all** condition and the **extents** condition must be met.

The value of a system token is defined by the following formula:

$$\begin{aligned} Value(sys,ref) &= Element\_Value(sys,ref) \cdot Arguments\_Value(sys,ref) \\ &\quad - W_{FA} \cdot Element\_Value(sys) \cdot Arguments\_Value(sys,\overline{ref}) \end{aligned}$$

In this expression for the *Value* of a system token, *Arguments_Value*(*sys*,*ref*) and *Arguments_Value*(*sys*, $\overline{ref}$) measure the value of system arguments that correspond and that don't correspond to reference arguments, respectively.

Element_Value is a function of the attributes of the system token and, if mapped, how well they match those of the corresponding reference token. The inherent value of a token is defined as the product of the token's attribute value parameters, *AttrValue*, for the attributes ***type*** and ***modality***. This inherent value is reduced for any attribute errors (i.e., for any difference between the values of system and reference attributes), using the error weighting parameters, $\{W_{err\text{-}attribute}\}$. If a system token is unmapped, then the value of that token is weighted by a false alarm penalty, W_{FA}.

$$Element\_Value(sys,ref) = \min\left(\begin{array}{l} \prod_{\substack{attribute= \\ type,modality}} AttrValue(attribute_{sys}), \\ \prod_{\substack{attribute= \\ type,modality}} AttrValue(attribute_{ref}) \end{array} \right) \cdot \prod_{\substack{attribute= \\ type,subtype,modality, \\ genericity,polarity,tense}} W_{err-attribute}$$

$$Element\_Value(sys) = \prod_{\substack{attribute= \\ type,modality}} AttrValue(attribute_{sys})$$

Arguments_Value is a function of the mutual argument value (*MAV*) between the arguments of the system token and, if mapped, those of the corresponding reference token. An argument's *MAV*, if mapped, is equal to the mapped value of the elements serving as arguments, *Value*(arg_{sys},arg_{ref}), but reduced in value if the system's argument role is in error.

$$MAV(arg_{sys},arg_{ref}) = Value(arg_{sys},arg_{ref}) \cdot W_{err-role} \quad \text{if } Mentions\_Value(arg_{sys},arg_{ref}) > 0$$

$$MAV(arg_{sys}) = Value(arg_{sys},arg_{sys})$$

There are several requirements that must be satisfied in order for a reference argument to be considered to be in correspondence to a system argument. These requirements for correspondence are listed in Table 21.

Table 21 Conditions required for correspondence between system and reference event arguments

Condition	Requirement
Always	The reference argument must be mappable to the system argument. That is, they must have at least one mention in correspondence.
If the "**mapped**" argument option is invoked	The reference argument must correspond to the system argument. That is, they must be mapped to each other at the argument level.

For each pairing of a system event token with a reference event token, an optimum correspondence between system arguments and reference arguments that maximizes *Arguments_Value* is determined and used. This optimum mapping is constrained to be a one-to-one mapping between system and reference arguments.

Arguments_Value is computed using the following formula:

$$Arguments\_Value(sys, ref) = \sum_{\substack{all\ arg_{sys}\ with\ a \\ corresponding\ arg_{ref}}} \left(\sum_{\substack{all\ docs\ that \\ mention\ the\ relation}} MAV_{doc}(arg_{sys}, arg_{ref}) \right)$$

$$Arguments\_Value(sys, \overline{ref}) = \sum_{\substack{all\ arg_{sys}\ with\ \mathbf{no} \\ corresponding\ arg_{ref}}} \left(\sum_{\substack{all\ docs\ that \\ mention\ the\ relation}} MAV_{doc}(arg_{sys}, arg_{sys}) \right)$$

The current default scoring parameters for VDR are given in

Table 22 Default parameters for scoring VDR performance

***Element_Value* parameters**			***Arguments_Value* parameters**	
Event mapping requirements (Table 20) = **"arguments"**			**"mapped"** *arguments optional requirement* **NOT** *invoked* (Table 21)	
W_{FA} = 0.75				
attribute	***AttrValue***	$W_{err\text{-}attribute}$	$W_{err\text{-}role}$	0.75
Type	1.00 for all types	0.50		
Subtype	(not applicable)	0.90		
Modality	1.00 for all types	0.75		
Genericity	(not applicable)	1.00		
Polarity	(not applicable)	1.00		
Tense	(not applicable)	1.00		